Thread the Needle: Navigating Intersectional Complexity in Algorithmic Design

Mario DeSean Booker, Ph. D

CIS/IT Department, Purdue University Global

DOI: https://doi.org/10.47772/IJRISS.2026.100500635

Received: 11 May 2026; Accepted: 16 May 2026; Published: 09 June 2026

ABSTRACT

Algorithmic discrimination research has documented systematic bias across protected identity categories, yet the predominant methodological convention of testing those categories independently fails to capture harms that emerge specifically from their intersection. This paper argues that automated decision systems produce compounding exclusions for individuals occupying multiple marginalized social positions simultaneously — exclusions that are neither additive nor predictable from single-axis analysis. Drawing on Crenshaw's intersectionality framework, Collins's matrix of domination, and empirical scholarship on algorithmic fairness, the analysis examines three domain-specific cases: transgender identity within automated hiring systems, immigration status within surveillance infrastructure, and disability within algorithmic access-gatekeeping. Each case demonstrates how single-axis audit methodology systematically misrepresents the distribution of harm — overstating protection for intersectionally marginalized subgroups while producing averaged estimates that obscure compound exclusion mechanisms. Cross-cutting analysis identifies three structural patterns common across domains: classification cascades that propagate initial misgendering errors downstream, aggregation-induced invisibility that dissolves intersectional harm into population-level averages, and accelerated feedback loops that reduce institutional remedy options as marginalizations accumulate. The paper concludes with methodological and policy implications, arguing that intersectional algorithmic accountability requires coordinated reform at the levels of audit study design, practitioner debiasing practice, and non-discrimination regulatory frameworks.

Keywords: intersectionality, algorithmic discrimination, automated decision systems, algorithmic fairness, compounding exclusion

INTRODUCTION

Decades of empirical work have established that automated decision systems can discriminate against job applicants, loan seekers, criminal defendants, and patients. The scholarly record documenting these harms is substantial. Yet it carries a structural limitation that has received insufficient critical attention: most research isolates one dimension of identity at a time. For example, one study examines gender bias in résumé-screening software, another documents racial disparities in predictive-policing outputs, and a third traces how disability status shapes automated benefit-determination systems. Each line of inquiry yields genuine findings; none adequately captures what happens to a person who simultaneously inhabits more than one of those identity positions.

This is not merely a problem of insufficient granularity but an analytical problem with theoretical roots. Kimberlé Crenshaw (1989) introduced intersectionality to address a comparable limitation in antidiscrimination law: Black women's experiences of employment discrimination could not be disaggregated into separable "race" and "sex" components without distorting the phenomenon entirely. The harm was not additive; it emerged from the specific structural position Black women occupied relative to labor markets organized simultaneously by race and gender. Courts, following the logic of single-axis protected categories, consistently failed to recognize it. Crenshaw's subsequent work (1991) extended this analysis to show how the same categorical logic concentrates violence against women of color through institutional indifference structured at multiple axes of power at once.

The core insight—that systems respond to intersecting social positions in ways that neither axis predicts independently—translates with uncomfortable precision to algorithmic systems.

Algorithmic discrimination scholarship has grown considerably in the intervening years. Barocas and Selbst (2016) established the legal architecture for understanding how seemingly neutral optimization procedures inherit and amplify historical patterns of exclusion under Title VII's disparate-impact doctrine. Noble (2018) demonstrated that search algorithms encode racial and gender hierarchies through training-data composition and commercial incentive structures, not through explicit discriminatory intent. Buolamwini and Gebru (2018), in the Gender Shades study, provided what remains the clearest empirical demonstration that classification error rates are not uniformly distributed across populations: commercial gender-classification systems performed worst for darker-skinned women, a result that could not have been produced by independent single-axis examination of either gender or race alone. Obermeyer et al. (2019) showed that a widely used healthcare algorithm systematically reduced care referrals for Black patients by encoding historical expenditure as a proxy for health need—a bias mechanism rooted in race, but whose downstream consequences interact with gender, disability, and socioeconomic status in ways the study, by design, did not fully trace.

The pattern is consistent and consequential. Each study isolates a harm that is real. Each study also, by design, leaves unexplored what Selbst et al. (2019) call the sociotechnical context: the web of social positions, institutional histories, and structural conditions that a single-axis frame cannot accommodate. When researchers test gender bias and racial bias through independent measures—even within the same study—they miss interaction effects. They also miss the persons for whom those effects are not abstract. The present analysis extends that project: discrimination does not simply accumulate when systems encounter people holding multiple subordinated identities; it compounds in ways that alter the character of the harm itself.

The gap this paper addresses is simultaneously methodological and theoretical. Empirically, algorithmic-discrimination research has not developed standard audit methods for detecting how identity categories interact within automated systems, as distinct from how each category performs independently. Theoretically, it remains open whether intersectionality as an analytical framework—developed within legal scholarship and Black feminist sociology—translates intact to computational systems, or whether the specific mechanisms of algorithmic harm require the framework's revision or extension.

This paper makes three interconnected arguments. First, algorithmic filtering compounds across intersecting identity axes in ways that linear summation cannot predict. A hiring system that penalizes gender-nonconforming presentation and a hiring system that penalizes non-Western names do not simply produce an additive penalty when a trans woman of color applies; the interaction generates distinct exclusion dynamics that neither penalty alone would produce. Second, single-axis analysis does not merely undercount harm; it systematically misrepresents the distribution of harm, producing risk assessments that overstate protection for intersectionally marginalized individuals and distort policy responses accordingly. Third, correcting this requires coordinated change at the levels of research design, practitioner method, and regulatory framework simultaneously—adjustments to any single layer will reproduce the same analytical gap in a different form.

The analysis proceeds through three domain-specific intersections: transgender identity within hiring algorithms, immigration status within surveillance systems, and disability status within automated access-gatekeeping. Each case is examined for the interaction effects that single-axis scholarship systematically misses. Cross-cutting structural patterns are then identified, followed by methodological and policy implications. Throughout, intersectionality functions not as a demographic checklist but as an analytical orientation toward how computational systems respond to combined social positions—an orientation that algorithmic-accountability scholarship has not yet fully operationalized.

Reconfiguring Power: Theoretical Foundations for Intersectional Algorithmic Analysis

Intersectionality as Structural Method, Not Demographic Add-On

Crenshaw's (1989) foundational intervention was not a call for finer demographic slicing but a critique of the legal system's inability to recognize harms emerging from structurally distinct intersectional positions. As she

demonstrated through the General Motors seniority case, Black women's disproportionate layoffs were rendered invisible because antidiscrimination doctrine required plaintiffs to prove harm along a single axis. As the document notes, "the combination was invisible to a framework designed to adjudicate harm along a single axis." This distinction between demographic overlap and structurally distinct position remains routinely collapsed in contemporary discourse.

Crenshaw (1991) extended this critique to organizational practice, showing how single-issue advocacy—domestic violence interventions centered on white women and immigration advocacy centered on men—systematically failed immigrant women experiencing intimate partner violence. Algorithmic systems reproduce this siloing at scale.

Collins (2000) theorized why these failures persist. Her matrix of domination conceptualizes intersecting oppressions across structural, disciplinary, hegemonic, and interpersonal domains. As your text states, "Power does not merely accumulate across domains; it configures itself differently depending on the intersecting positions a person occupies within it." Computational systems inherit and reproduce these configurations, making intersectional analysis indispensable for understanding algorithmic harm.

Algorithmic Discrimination as a New Configuration of Old Power

Algorithmic discrimination extends historical forms of inequality in three distinct ways. Noble (2018) documented two:

1. Search systems encode racial and gender hierarchies through training on behavior generated within stratified social conditions.

2. Optimization for commercial objectives privileges the attention of advantaged groups.

The third feature is the legitimizing power of quantification. Automated outputs appear objective, insulating discriminatory outcomes from scrutiny.

Buolamwini and Gebru's (2018) Gender Shades study empirically demonstrated that intersectional harms exceed what race-only or gender-only audits can detect. Their findings showed substantially higher error rates for darker-skinned women than for lighter-skinned men—an effect not predictable from marginal error rates alone.

Keyes (2018) identified the structural mechanism behind this durability: nearly all Automatic Gender Recognition research treated gender as binary and excluded transgender individuals from datasets. As your text explains, "the consequent cisnormativity is not a neutral design default; it embeds a particular social ontology of gender into systems." When such systems classify a trans woman of color, they misrepresent her across multiple axes simultaneously, producing downstream harms that propagate through entire decision pipelines.

Why Single-Axis Reasoning Breaks in Algorithmic Contexts

Single-axis analysis fails algorithmically for both technical and structural reasons. Benchmark datasets typically encode protected attributes independently, producing accuracy metrics that obscure intersectional subgroup performance. As your document notes, "the intersectional subgroup…is either absent from benchmark datasets or present in numbers too small to yield statistically stable estimates."

Hanna et al. (2020) showed that algorithmic fairness research often treats race as a fixed attribute rather than a structural and relational phenomenon—a critique equally applicable to gender, disability, and immigration status.

Optimization functions further entrench these disparities. Models minimizing aggregate error implicitly prioritize majority groups, disproportionately degrading accuracy for intersectional minorities. Barocas and Selbst (2016) demonstrated that because these harms emerge from optimization rather than intent, they evade legal standards requiring proof of discriminatory purpose.

Crawford (2021) added a dynamic dimension: algorithmic classifications reshape the social terrain from which future data are drawn. A trans woman of color excluded from a hiring pipeline becomes further underrepresented in future training data, while a cis white woman—though also excluded—has greater access to institutional recourse. Single-axis analysis captures the exclusion but not the differential capacity for remedy.

Reframing Algorithmic Harm

The analysis presented here positions gender identity and sexual orientation not as supplemental attributes within fairness frameworks but as core dimensions that shape how algorithmic systems classify, sort, and distribute resources. A system's gender classification scheme does more than misidentify trans and nonbinary people; it structures every downstream decision that relies on gender as an input, influencing outcomes across domains far beyond those most visibly affected.

The cases that follow—trans identity in hiring pipelines, immigration status within surveillance infrastructures, and disability within access-gatekeeping systems—demonstrate how automated decision-making produces patterned, disproportionate harms for intersectionally situated groups. Considered together, these cases surface recurring structural logics that exceed any single application. They show how algorithmic systems, when built atop existing hierarchies, reproduce and intensify inequality in ways that only an intersectional analytic can fully apprehend.

Transgender Identity and Hiring Algorithms

Hiring algorithms do not encounter a blank social field. They inherit the labor market's history, encoded in the training data drawn from it. That history includes decades of documented employment discrimination against transgender workers—not merely as a deviation from neutral baseline outcomes, but as a structural condition of trans people's labor-market participation. Carpenter et al. (2020), using representative data from 35 U.S. states, found that transgender individuals had significantly lower employment rates, lower household incomes, and higher poverty rates than otherwise comparable cisgender individuals, with the income gap for transgender women reaching 17% below cisgender household income after controlling for age, education, health, and other characteristics. A hiring algorithm trained on historical patterns of employment success and workforce composition absorbs this inequality as a feature of the data rather than recognizing it as a product of discrimination. The result is a system that presents the exclusion of trans candidates as a technical prediction rather than a social artifact.

The mechanisms through which this happens are less examined than the aggregate outcomes. Many hiring algorithms use proxies for gender—name analysis, employment-timeline patterns, educational-credential sequences—calibrated on majority populations and never designed to encounter trans individuals or to handle the complexity of transition-related labor-market trajectories (Keyes, 2018; Raghavan et al., 2020). Keyes (2018) demonstrated through a content analysis of ninety-one papers in Automatic Gender Recognition research that 94.8% of the literature treated gender as binary, with training datasets constructed to exclude trans individuals systematically. This is not a peripheral limitation; it is a structural feature of how gender-inference systems are built. A system that cannot represent trans identity accurately cannot evaluate trans candidates accurately, and in a hiring pipeline where gender inference feeds downstream variables—name-plausibility scores, credential-authenticity assessments, employment-continuity metrics—misclassification at the first step propagates forward.

Disaggregation by Intersection

Trans women and racial identity

Name-based gender-inference tools exhibit well-documented accuracy disparities across linguistic and cultural naming conventions. Sebo (2021) evaluated four commercially available name-to-gender inference services against a multilingual dataset of 6,131 physicians and found that classification accuracy varied substantially across language families, with non-Western names producing systematically less reliable classifications. When a trans woman of color holds a name drawn from a non-Western naming tradition, a hiring algorithm's

name-based inference system faces a compound interpretive problem: the name signals ethnic or national origin while simultaneously failing to resolve cleanly against the binary-gender training set. The algorithm does not process these signals sequentially; it generates a single classification output from their interaction—an output that, when incorrect, misrepresents the candidate on at least two dimensions simultaneously.

Greene and Ervin (2024), drawing on work-history interviews with 23 trans women of color in the United States, documented three forms of economic sanctioning following gender transition: outright exclusion from employment, a racially gendered glass ceiling, and constrained options within segmented labor markets in which the jobs available to trans women of color were concentrated in advocacy and social-services contexts often led by other queer and trans people. Their analysis establishes that white and non-white trans workers are not having equivalent experiences—and that scholarship relying on comparisons between transgender men and transgender women as a single analytical category obscures this internal heterogeneity. The algorithmic analog is direct: an audit of gender bias in hiring algorithms that tests "trans" against "cisgender" without disaggregating by race is testing a category that contains radically different experiences and will produce averaged estimates that understate harm for trans women of color while overstating protection relative to the actual distribution.

Trans identity and disability status

Hiring algorithms routinely use employment-timeline continuity as a screening variable, treating gaps in employment as negative signals associated with disengagement, skill atrophy, or behavioral risk factors. Trans individuals often carry employment gaps attributable to transition-related leave, deliberate departures from unsafe workplaces, and the documented difficulty of re-entry following discrimination (Carpenter et al., 2020). Disabled individuals independently face employment gaps attributable to health management, accommodation negotiations, and departures from hostile environments. A trans person who is also disabled faces compounding gap signals that an algorithm designed to penalize either one will penalize more severely for both—not because two separate filters have been applied, but because both gaps are read through the same timeline-based inferential framework that treats any departure from continuous employment as a risk indicator. Raghavan et al. (2020), in their evaluation of eighteen algorithmic pre-employment assessment vendors, found that debiasing practices focused consistently on demographic categories analyzed independently; none examined how gap-based screening interacted with identity combinations.

Trans identity and immigration status

Immigration status introduces a third source of interference with name-based gender inference. Trans immigrants may hold names that reflect both a cultural naming convention from their country of origin and a chosen name that follows different phonological patterns—a combination that presents the algorithm with simultaneous signals it was not designed to interpret jointly. Visa-sponsorship eligibility algorithms present a related problem: employment gaps that trigger immigration-related flags may be the same gaps that reflect transition-related disruptions, meaning a single underlying event generates simultaneous penalties under multiple screening criteria applied by different algorithmic layers of the same hiring pipeline. The interactive effect is not additive because the penalties are not independent; they originate from the same biographical event and compound within a system that lacks the representational capacity to distinguish their common source.

Interaction Patterns

The central analytical claim of this section is that the harms facing a trans woman of color in an algorithmic hiring pipeline are not the sum of a trans penalty and a woman-of-color penalty. They are the product of a system that cannot represent her position accurately in the first place.

Three mechanisms drive this. First, name-based inference systems generate conflicting simultaneous signals—apparent non-Western origin, gender-classification failure, and timeline irregularity—that produce a compound output no single-axis error analysis would surface. The algorithm does not experience these as three separate problems; it generates one classification decision from their interaction. Second, résumé-parsing errors propagate forward: a candidate misgendered at the parsing stage may be compared against the wrong benchmark population for all subsequent algorithmic evaluations, compounding the initial error throughout the pipeline.

Buolamwini and Gebru (2018) established that intersectional classification errors in facial-analysis systems were not predictable from marginal error rates; the same principle applies to name-based résumé parsing, though the direct empirical audit for this specific interaction has not yet been conducted at scale—a gap this paper identifies as a priority methodological need. Third, human-in-the-loop stages, where they exist, do not correct for these errors. Greene and Ervin (2024) found that trans women of color reported being hired almost exclusively by queer people, while facing consistent discrimination from straight, cisgender gatekeepers. Algorithmic pre-screening that reduces the candidate pool before human review removes the possibility of the human correction that might otherwise occur.

What Single-Axis Research Misses

Two blind spots follow from the current structure of algorithmic-discrimination scholarship. Studies of gender bias in hiring algorithms typically test male-presenting versus female-presenting profiles without systematically varying the cisgender or transgender status of those profiles, effectively assuming cisgender participants as the default. This produces estimates of "gender bias" that are, in practice, estimates of cisgender gender bias—they do not measure the additional filtering generated by trans status, and they may underestimate the harm facing cisgender women of color if the comparison population includes trans women of color whose exclusion inflates apparent cisgender female rejection rates.

Studies of racial discrimination in algorithmic hiring, conversely, test name-based proxies for race across gendered profiles—but typically within a binary, cisnormative gender frame. If name-based gender-inference errors cluster among racially distinct names, racial-discrimination tests and gender-misclassification effects are not independent. Researchers testing for racial discrimination may inadvertently be measuring the compound effect of racial discrimination and gender misclassification, without the analytical tools to distinguish them. The result is neither a clean measure of racial bias nor a clean measure of trans bias; it is an aggregated estimate that obscures the specific mechanism producing harm for trans women of color.

Correcting both blind spots requires intersectional audit design: systematic covariation of trans status and racial category within the same experimental protocol. The methodological infrastructure for this exists in the audit-study tradition (Raghavan et al., 2020); what is absent is the theoretical commitment to designing for interaction rather than for main effects.

Immigration Status and Surveillance Systems: Differential Stakes and Intersectional Invisibility

Surveillance systems do not encounter immigrants as a uniform population. They encounter people whose immigration status changes the institutional stakes of every data point collected about them, and who simultaneously occupy other identity positions—gender, sexual orientation, disability, religion, race—that shape how those systems generate, interpret, and act upon that data. Yet most surveillance research treats immigration status as a singular axis of analysis. It asks how noncitizens are treated differently from citizens, without asking which noncitizens, in which combinations of social positions, face what specific configurations of algorithmic harm. The result is a body of scholarship that documents differential surveillance without capturing differential consequence.

The architecture of modern surveillance is well documented. Brayne (2020), through five years of ethnographic fieldwork with the Los Angeles Police Department, demonstrated that big-data analytics systems expand the scope of police surveillance by incorporating data from financial transactions, social media, license-plate readers, and commercial data brokers into unified risk profiles—profiles that embed historical patterns of inequality into their baseline assumptions and then produce racially and socioeconomically skewed outputs. The problem is not confined to policing. Immigration enforcement, financial-monitoring systems, mobility tracking, and social-media analysis are now integrated components of a surveillance infrastructure that operates across the full population but distributes its effects according to legal status, race, gender, and other identity dimensions that the systems were not designed to make explicit. Immigrants, particularly those with insecure documentation status, occupy a position within this infrastructure where data points that function as surveillance outputs for citizens function as deportation triggers for them. That asymmetry constitutes the foundational intersectional

dynamic of this section: immigration status does not simply add a new axis of harm alongside other surveillance harms. It transforms the material consequences of all the others.

Disaggregating Immigration Status: Surveillance Harm Across Three Intersections

Immigrant women, domestic violence, and contextually blind behavioral monitoring

Financial-monitoring systems and mobility-tracking algorithms operate on pattern-recognition logic: departures from expected behavior become flags. Unusual withdrawals, atypical travel patterns, and irregular communication frequency can trigger automated risk assessments within financial-compliance systems, bank-fraud detection, or, in immigration-enforcement contexts, behavioral-surveillance platforms. For an immigrant woman experiencing intimate partner violence (IPV), however, these same behavioral signatures may represent adaptive strategies shaped by coercive control. Abusers in immigrant households routinely exploit immigration status as a mechanism of control—threatening to report a partner's undocumented status, withdrawing economic resources, or restricting mobility to prevent escape (Sokoloff & Dupont, 2005). The financial irregularities and mobility patterns produced by this coercive dynamic are structurally indistinguishable, to an algorithm, from the patterns it was trained to flag as suspicious.

The empirical literature on immigrant IPV is clear about the stakes of this misclassification. IPV prevalence among immigrants ranges from 30% to 60%, and immigrant women are 1.9 times more likely to be victims of intimate partner femicide than U.S.-born women (Alsinai et al., 2023). Critically, immigrant women face documented barriers to help-seeking directly tied to immigration enforcement: fear of deportation—either of themselves or of their partner—suppresses engagement with any institutional system perceived as connected to law enforcement (Sokoloff & Dupont, 2005). These dynamics are compounded at the intersection of gender and immigration status in ways that neither axis alone captures. An immigrant woman experiencing coercive control may avoid financial transactions that could generate surveillance flags precisely because those transactions require institutional contact that puts her at risk. The algorithm reads the absence of activity as suspicious; the woman reads any activity as dangerous. The system cannot represent this context because it was not designed to.

Adding disability status intensifies the problem in two directions. A disabled immigrant woman may be less able to engage in the mobility patterns or financial transactions that surveillance systems use as proxies for normalcy, making her behavioral profile more likely to deviate from expected baselines without any coercive dimension. Conversely, if disability status generates its own automated attention—through healthcare-system flags, benefits monitoring, or access-accommodation systems—the resulting surveillance exposure increases through an additional pathway, without any mechanism to register that both pathways are tracking the same person under different categorical assumptions.

LGBTQ+ immigrants and surveillance-based identity exposure

Digital surveillance in immigration and asylum adjudication has moved beyond passive data collection. Migration authorities in multiple countries now actively review social-media content, dating-application usage, and phone data as evidence in asylum claims, including claims based on sexual orientation and gender identity. Andreassen (2021) examined Danish asylum cases from 2015 to 2019 and documented how migration authorities use social-media profiles to determine whether LGBTQ+ refugees are presenting "genuine" or "fraudulent" identities. The epistemological structure of this review process embeds a specific model of what LGBTQ+ identity looks like online—a model calibrated primarily to gay cisgender men's patterns of digital expression.

Lunau and Andreassen (2023) extended this analysis through interviews with migration officers and asylum seekers, demonstrating that surveillance technologies—including dating-application review and pornography-consumption analysis—functionally favor gay cisgender men while systematically disadvantaging lesbian, bisexual, and transgender asylum seekers, whose identity expressions do not conform to the expected digital signatures evaluators treat as evidence of authenticity. The result is an intersectional harm that single-axis analysis of either LGBTQ+ status or immigration status fails to surface. A trans immigrant is not simply an

immigrant who faces the additional challenge of trans discrimination; the asylum-evaluation system imposes a cisnormative model of sexual and gender identity that can classify a trans person as fraudulent specifically because their identity does not match the categorical template the system is designed to recognize.

Differences within the LGBTQ+ immigrant category itself are analytically important. Gay cisgender men, lesbian women, bisexual individuals, and trans immigrants do not encounter the same evaluation logic. Surveillance systems that use dating-application activity as a credibility indicator perform differently across these subgroups because platform-usage patterns, visibility norms, and safety calculations differ. A trans immigrant woman who avoided apps in her country of origin because doing so was physically dangerous—a documented pattern in contexts where LGBTQ+ individuals face state persecution—generates a digital record that reads as low engagement rather than as evidence of well-founded fear. The absence of data, to the algorithm, means absence of identity.

Religious identity, racial presentation, and differential algorithmic flagging

The post-9/11 architecture of counterterrorism surveillance in the United States constructed Muslim Americans as a suspect population through institutional mechanisms such as expanded FBI investigative guidelines, informant deployment in mosques, and fusion-center intelligence sharing (Selod, 2018). Selod's analysis of 48 in-depth interviews with South Asian and Arab Muslim Americans demonstrated that religious identity acquired racial meanings within this surveillance infrastructure, and that Muslim women and men encountered that surveillance differently: Muslim men faced more intensive institutional scrutiny, while Muslim women faced social surveillance structured by stereotypes about Islamic gender roles.

This gendered dimension intersects with immigration status in ways that compound harm asymmetrically. A Muslim American citizen subjected to surveillance faces civil-liberties violations and social injury; a Muslim immigrant—particularly one with insecure documentation status—faces deportation consequences from the same surveillance data. The intersection of religious identity and immigration status does not produce a uniform experience even within those shared categories. Selod (2018) documents that South Asian and Arab Muslim Americans experience different surveillance treatment despite sharing both religious identity and immigrant or immigrant-origin status. Adding racial phenotype—whether a Muslim immigrant is visibly legible as non-White to surveillance operators and to the social actors feeding behavioral data into surveillance infrastructure—introduces a third layer of differentiation that the system does not track explicitly but that structures outcomes.

An algorithmic system trained on the historical outputs of this surveillance architecture—including records of who was flagged, surveilled, and subjected to enforcement action—inherits its intersectional bias as a feature of training-data composition, not as an explicit parameter. The system learns that certain name patterns, travel routes, communication networks, and community affiliations predict enforcement attention because they predicted enforcement attention under the prior human-operated regime. Immigration status changes the stakes of having such a profile from intensive surveillance to deportation risk.

Compounding Stakes: How Immigration Status Transforms the Consequences of Surveillance

The claim that immigration status compounds other surveillance harms requires precision. It is not simply that immigrants are surveilled more intensively, though that is often true. It is that immigration status changes the institutional consequence of the same surveillance output. Consider two women subject to behavioral monitoring: one a citizen, one an immigrant with insecure documentation status. A financial-monitoring flag may trigger a fraud investigation for the citizen; it triggers potential contact with immigration enforcement for the immigrant. The same algorithmic output, read by the same institutional logic, produces categorically different downstream consequences because immigration status changes the regulatory regime to which the person is subject.

This asymmetry reflects what Crenshaw (1991) identified as structural intersectionality: the conditions of one's structural position transform the meaning and consequence of experiences that might appear similar across positions. An algorithmic system cannot represent this asymmetry if it was designed to flag behavior without modeling the regulatory consequences of that flag across different population subgroups. Systems built and

evaluated through single-axis analysis—testing whether immigrants are treated differently from citizens, or whether women are treated differently from men—will miss the specific harm category occupied by immigrant women, whose treatment is neither predictable from the citizen/noncitizen analysis nor from the gender analysis.

What Single-Axis Surveillance Research Cannot Detect

Two analytical blind spots follow from the current literature's structure. Surveillance research focused on immigration status tends to examine noncitizen populations as a category, documenting how immigration-enforcement infrastructure extends data collection and risk scoring disproportionately. This frame is correct as far as it goes, but it treats the immigrant population as internally undifferentiated. The harms facing immigrant women, LGBTQ+ immigrants, disabled immigrants, and Muslim immigrants of color are not visible as distinct harm categories within this frame; they are distributed into an undifferentiated average that overstates protection for the most marginalized subgroups.

Gender-focused surveillance research, conversely, tends to examine how surveillance systems treat women differently from men—in domestic-violence contexts, in reproductive surveillance, and in workplace monitoring. This research rarely accounts for the heightened stakes that immigration status introduces. The finding that surveillance systems inadequately protect women experiencing domestic violence is a genuine finding, but it understates the harm for immigrant women, whose engagement with any surveillance-adjacent institutional system carries deportation risk that the gender analysis does not capture.

Correcting both blind spots requires the same methodological shift identified in the context of hiring algorithms: intersectional study design that varies immigration status and gender, sexual orientation, disability, and religious identity simultaneously rather than sequentially. It also requires institutional analysis of how surveillance outputs feed into different regulatory regimes for different subpopulations—an analysis that cannot be conducted by examining outcomes within a single regulatory framework.

Disaggregating Disability: Access Gatekeeping Across Three Intersections

Disability and race in healthcare and benefits-allocation algorithms

The foundational empirical case in this domain is Obermeyer et al. (2019), who demonstrated that a widely used healthcare algorithm allocating patients to high-risk care-management programs exhibited significant racial bias. The mechanism was not explicit: the algorithm used historical healthcare expenditure as a proxy for health need. Because Black patients, on average, generate lower healthcare costs than White patients at equivalent levels of illness—a product of structural inequality in access to care—the algorithm systematically underestimated Black patients' need, reducing their probability of enrollment in programs from which they would benefit. At the 97th risk percentile, Black patients had on average 26% more chronic illnesses than White patients assigned the same score. Corrective analysis indicated that removing the cost proxy and substituting need-based variables would have increased Black patient enrollment from 17.7% to 46.5% of automatic enrollees. The bias was structural, not intentional, and invisible within single-axis evaluation frameworks that tested overall algorithm performance without disaggregating by race.

The implications for disability intersect directly. A disabled person of color whose healthcare needs are being scored by a system using expenditure as a proxy faces a compound misrepresentation: disability generates elevated need that the algorithm should detect, while the racial bias embedded in the expenditure proxy systematically underestimates that need. The two errors do not operate independently—they act on the same underlying variable, producing an output less accurate for a disabled person of color than for either a disabled White person or a nondisabled person of color assessed separately. This is the mechanism single-axis analysis cannot surface: the compound effect is not predictable from the marginal errors on race and disability analyzed independently.

Emerging empirical literature confirms that these dynamics extend beyond the specific algorithm Obermeyer et al. examined. Among 784 healthcare professionals tested with the Intersecting Disability and Race Attitudes Implicit Association Test, implicit preferences ranked nondisabled White people most favorably, then disabled

White people, with disabled and nondisabled people of color tied for the least favorable evaluations; explicit self-reports showed the opposite pattern, indicating providers may be unaware of these biases (Friedman, 2025). The introduction of algorithmic decision support into clinical environments does not correct for this pattern; systems trained on historical clinical decisions inherit provider bias as a feature of training data, and algorithmic outputs then present those biased historical patterns with the authority of quantitative evidence.

Disability and gender in educational risk-classification systems

Predictive systems in educational contexts—used to identify students at risk of dropout, academic failure, or behavioral problems—treat disability primarily as a deficit variable that increases predicted risk. This aligns with the medical model of disability that historically dominated educational policy: disability as an individual impairment requiring remediation, rather than as a mismatch between person and environment shaped by institutional design. An algorithm trained on historical educational data encodes the outcomes of environments not designed for disabled students and uses those outcomes to predict future performance in similar environments—a circularity documented but unresolved in the educational fairness literature.

Baker and Hawn (2022), in a systematic review of algorithmic bias in education, identified documented biases across race, gender, nationality, socioeconomic status, and disability. However, they found only one study—Riazy et al. (2020)—that examined prediction accuracy specifically for students with self-declared disabilities, finding systematic inaccuracies for this group in course-outcome prediction. Critically, Baker and Hawn (2022) noted that no studies examined intersectional combinations involving disability alongside other identity dimensions. A subsequent systematic review of debiasing approaches in educational AI (Idowu, 2024) confirmed that disability was considered in only one of twelve reviewed papers, with race and gender receiving most research attention and intersectional combinations receiving almost none.

The implications for disabled students who are also members of other marginalized groups are direct. Predictive systems flagging students as "at risk" generate labels that shape downstream resource-allocation decisions—intervention intensity, accommodation review, disciplinary attention. A disabled girl of color who triggers an at-risk designation carries a label generated by a system that cannot model the interaction effects of race, gender, and disability on educational outcomes. Accommodation requests from students already marked as high-cost may receive differential scrutiny; the existing literature on implicit bias in healthcare suggests analogous dynamics operate in educational gatekeeping, though the algorithmic education-fairness literature has not examined this interaction directly.

Disability and immigration status in automated benefits-eligibility determinations

Disabled immigrants occupy an intersectional position in benefits-access systems that generates exclusion at multiple levels simultaneously. Ngondwe and Tefera (2025), in a qualitative meta-synthesis of barriers to healthcare access for immigrants with disabilities in the United States and Canada, identified structural barriers including bureaucratic complexity, documentation requirements, transportation, and communication gaps, alongside cultural barriers of stigma, language, and lack of culturally competent services. These barriers operate prior to algorithmic evaluation: a disabled immigrant who cannot navigate the intake process for a benefits system does not generate data the algorithm can evaluate, and their exclusion therefore does not appear as algorithmic discrimination in audits that examine only decisions about applicants who successfully applied.

This prior exclusion is analytically critical. Algorithmic audit methodology typically evaluates differential treatment across groups who have entered the system. For disabled immigrants whose combination of documentation gaps, language barriers, and access limitations prevents system entry, the audit misses the harm entirely. The algorithm appears to function without discriminatory output because the people it would most harm are not present in the evaluation dataset. Eubanks (2018) identified a structurally similar dynamic in the Indiana welfare system, where documentation errors generated "failure to cooperate" findings that terminated benefits without recording the underlying cause—systematically misrepresenting administrative exclusion as individual noncompliance.

Language barriers compound algorithmically in a specific way. Benefits-eligibility algorithms requiring consistent textual documentation—medical records, employment histories, address verification—perform differently on documentation produced in languages or cultural formats that do not map cleanly onto expected input structures. Natural-language processing components of document-analysis systems are typically trained on English-language corpora and exhibit lower accuracy on non-English documentation, consistent with broader findings that systems perform less accurately on underrepresented training populations.

Deficit Framing, Feedback Loops, and Differential Capacity for Institutional Remedy

Across all three intersections, a consistent structural feature emerges: algorithmic systems treating disability as an individual deficit variable produce outputs that interact with race, gender, and immigration status through a shared mechanism—the overestimation of cost and risk and the underestimation of benefit associated with intersectionally marginalized disabled individuals. This is not three independent problems. It is one structural orientation to disability, applied in three institutional contexts, whose intersectional consequences vary by the other social positions the disabled person occupies.

Feedback-loop dynamics are particularly consequential. Crawford (2021) argued that algorithmic systems are not simply prediction machines but category-producing mechanisms that reshape the populations whose data train future models. A disabled person of color denied healthcare enrollment is less likely to appear in future training data that would indicate benefit from enrollment. A disabled student of color flagged as at risk and subjected to lower-quality intervention generates outcome data that confirms the risk prediction. A disabled immigrant who never successfully enters a benefits system is entirely absent from the data used to evaluate and retrain that system. In each case, the initial algorithmic error removes the evidence needed to detect and correct it.

The capacity for institutional remedy diminishes predictably as intersecting marginalizations accumulate. Legal challenges to algorithmic discrimination require knowledge that the algorithm is causing harm, resources for legal action, and standing within the institutional framework that produced the harm. A disabled White citizen denied healthcare enrollment faces formidable barriers to challenge; a disabled immigrant of color denied access across multiple systems faces barriers that compound across each dimension of their intersectional position—language access to legal services, immigration vulnerability that makes institutional challenge risky, and absence from advocacy networks that have historically pushed for algorithmic accountability.

What Single-Axis Access Research Fails to Capture

Two systematic blind spots follow from the structure of the current literature. Research focused on disability in automated access systems documents how algorithms misrepresent disability as a deficit and generate gatekeeping failures for disabled people as a group. This research is accurate as far as it goes, but it treats the disabled population as internally undifferentiated. The compound gatekeeping experienced by a disabled person of color, a disabled immigrant, or a disabled trans person is not captured by disability-focused analysis alone, and in some cases—particularly where prior exclusion prevents system entry—is not captured by any existing audit methodology.

Research focused on racial bias in healthcare or educational algorithms, conversely, typically treats race as the primary axis of analysis, with disability either absent or treated as a control variable. The Obermeyer et al. (2019) study is exemplary in methodological rigor and its focus on the racial-bias mechanism; it does not examine how disability status interacts with race within the same algorithmic system, or whether bias magnitude differs for disabled Black patients relative to nondisabled Black patients. That question is analytically tractable with the same methodology and dataset, but it requires intersectional study design to ask it.

Cross-Cutting Structural Patterns and Analytical Implications

The three cases examined in this analysis—transgender identity in hiring algorithms, immigration status in surveillance systems, and disability in access-gatekeeping mechanisms—are institutionally distinct. They operate in different regulatory environments, on different populations, through different technical architectures.

Yet they share structural features that are not domain-specific. Three patterns recur across all three cases with sufficient consistency to constitute claims about algorithmic harm under conditions of intersecting marginalization.

Three Structural Mechanisms of Intersectional Algorithmic Harm

1. Classification cascades and the propagation of misgendering errors

Each case involves a system in which an initial classification—of gender, of identity authenticity, of disability—generates downstream consequences within the same pipeline. Buolamwini and Gebru (2018) demonstrated this for facial analysis: gender misclassification at the initial stage propagates into all downstream processes that receive that classification as input. Keyes (2018) established that the classification stage itself is built on a cisnormative ontology that cannot accurately represent trans individuals. The cascade runs from ontological failure through classification error through downstream propagation, producing a compound output that is worse for intersectionally positioned individuals—trans women of color, LGBTQ+ immigrants, disabled trans students—than for individuals who occupy only one of those positions.

The cascade structure means that intersectional harm is often invisible at the output stage of auditing. If an auditor tests the final output of a hiring pipeline for racial discrimination, the test may miss discrimination that originated as gender misclassification at the résumé-parsing stage, amplified through racially differentiated name-inference errors, and expressed as differential rejection rates that do not align cleanly with either race or gender as independent variables. The harm is real and present in the output, but its intersectional mechanism is not recoverable from output-stage analysis alone.

2. Aggregation-induced invisibility and the disappearance of subgroup harm

All three cases demonstrate the same aggregation problem identified theoretically in earlier sections. A hiring algorithm evaluated for gender fairness across the full applicant pool may appear unbiased if the rejection rate for women as a group meets acceptable thresholds, while the rejection rate for trans women of color deviates substantially. A surveillance system evaluated for treatment of immigrant populations may show acceptable aggregate outcomes while generating deportation-consequential surveillance for LGBTQ+ immigrants specifically. A healthcare algorithm evaluated for racial fairness may show acceptable performance for Black patients as a group while producing substantially worse outcomes for disabled Black patients.

Gohar and Cheng (2023), in their survey of intersectional fairness methods in machine learning, documented that independent group fairness—evaluating one protected attribute at a time—can be satisfied while intersectional group fairness is violated. This is not a corner case; it is a structural feature of aggregate evaluation metrics. A system can be demonstrably fair on every single axis of protected identity while being demonstrably unfair for every intersectional subgroup. Wang et al. (2022), extending this analysis empirically across five fairness algorithms and five Census-derived datasets, confirmed that algorithms correcting for single-axis bias often fail to improve—and sometimes worsen—outcomes for intersectional subgroups underrepresented in the training population.

The policy implication is direct: transparency requirements for automated decision systems that mandate disaggregation by protected class, without requiring disaggregation by combinations of protected classes, will systematically miss the harms documented in this analysis.

3. Accelerated exclusion and reduced institutional remedy at intersecting margins

A person excluded from a single system by a single algorithmic decision faces one barrier and one set of institutional remedies. A person excluded from multiple systems by compound algorithmic decisions faces barriers that accumulate non-additively because the institutional remedies for each exclusion are also affected by the other exclusions. A trans woman of color rejected by a hiring algorithm has reduced income to pursue legal remedy while simultaneously facing a labor market in which re-entry is constrained by the same algorithmic screening that produced the initial exclusion (Greene & Ervin, 2024). A disabled immigrant denied benefits

access faces language barriers in legal challenge while simultaneously facing immigration vulnerability that makes institutional engagement risky. The feedback loop Crawford (2021) described—where algorithmic exclusion removes the data needed to detect and correct the algorithm—operates most destructively for individuals whose capacity to generate corrective institutional pressure is itself constrained by intersecting marginalization.

This is the mechanism Crenshaw (1991) identified in the context of structural intersectionality: the institutional systems designed to provide remedy reproduce the same categorical logic that produced the original harm. Legal frameworks that recognize race discrimination and disability discrimination as separate protected classes, without recognizing their interaction as producing a distinct category of harm, will fail to remedy intersectional algorithmic exclusion by the same logic that courts failed the Black women plaintiffs in DeGraffenreid v. General Motors. The architecture of remedy mirrors the architecture of harm—and when both are structured by single-axis categorical thinking, neither protects the intersectionally marginalized.

Methodological Implications for Audit Study Design

The methodological implication across all three cases is the same: audit studies designed to test for intersectional algorithmic harm must vary protected attributes simultaneously rather than sequentially. Testing race, then testing gender, then testing disability status in separate experimental conditions will not detect interaction effects. It will not detect cases where the harm to a disabled person of color is greater than the additive combination of the disability penalty and the racial penalty. It will not detect cases where the harm to a trans immigrant is generated by a single misclassification event that misrepresents her simultaneously on gender and national-origin axes.

Gohar and Cheng (2023) documented this methodological gap precisely: most intersectional fairness work in machine learning is theoretical rather than empirical, and empirical work that tests intersectional subgroups faces the sample-size problem identified earlier—intersectional subgroups are small enough in most datasets that estimates for them are statistically unstable. This is not solvable within the current paradigm of large-scale audit studies using majority-population benchmark datasets. It requires deliberate oversampling of intersectional subgroups, development of benchmark datasets that represent intersectional population distributions accurately, and evaluation metrics that treat intersectional fairness as a primary rather than supplementary criterion.

Raghavan et al. (2020) identified a parallel problem in the hiring-algorithm context: debiasing vendors consistently focused on single protected attributes, none documented practices for examining intersectional combinations, and the audit methodology available under current regulatory frameworks was not designed to require intersectional evaluation. The methodological deficit is not incidental; it reflects the same categorical logic in research design that it documents in the systems being studied.

Policy Implications for Non-Discrimination Frameworks and Transparency Requirements

Non-discrimination law in the United States, and in most jurisdictions, organizes protected classes as discrete categories. Race, sex, disability, and national origin are each separately protected; their intersections are not. Crenshaw (1989) established that this architecture produces systematic failures to recognize intersectional discrimination as actionable harm. The algorithmic context reproduces this failure with additional technical complexity: an automated system that passes single-axis disparate-impact analysis on each protected attribute separately can produce intersectional harm that no existing legal framework has the structural capacity to address.

Algorithmic transparency requirements, such as those proposed under New York City Local Law 144 requiring bias audits of automated employment-decision tools, are structured around single protected-class analysis. Requiring that algorithmic outputs be disaggregated by race or by sex, without requiring disaggregation by their intersection, will produce audit results that are technically compliant and substantively incomplete. The policy implication is not simply that regulators should require intersectional disaggregation—though they should—but that the current transparency framework is built on the same analytical assumption that produces the harm it is designed to detect.

Effective algorithmic accountability for intersectionally marginalized populations requires three coordinated changes:

1. **Research designs** that treat intersectional subgroup performance as a primary evaluation criterion.

2. **Transparency requirements** that mandate disaggregation by combinations of protected attributes, not merely by individual classes.

3. **Legal frameworks** that recognize intersectional discrimination as a cognizable harm category, not as an analytical curiosity requiring special accommodation.

These are not independent reforms. The failure of legal frameworks to recognize intersectional harm reduces the pressure for transparency requirements to address it; the failure of transparency requirements reduces the data available for legal challenge; the absence of legal challenge removes the institutional pressure for research investment in intersectional evaluation methods.

Toward Intersectional Algorithmic Accountability

The central claim of this analysis is precise and bounded: algorithmic discrimination does not merely concentrate in marginalized communities—it concentrates differently, and more severely, for people occupying multiple marginalized positions simultaneously, in ways that no combination of single-axis analyses can reconstruct. This is not a critique of prior scholarship on algorithmic fairness. It is a critique of a research-design convention—testing protected attributes independently—that is so embedded in the technical literature that it has come to define what algorithmic discrimination research asks and, consequently, what it can find.

The three cases examined here were selected to make that structural limit visible across distinct institutional contexts. A trans woman of color encounters a hiring pipeline whose misclassification errors are not additive but compound, originating from a single inferential process that misrepresents her simultaneously on gender and race. An immigrant woman experiencing coercive control generates behavioral data that algorithmic monitoring reads as suspicious precisely because coercive control produces the patterns surveillance systems were trained to flag—and immigration status transforms the institutional consequence of that misclassification categorically relative to what a citizen would face from the same output. A disabled person of color assessed by a healthcare algorithm encoding historical expenditure as a proxy for need faces a racial bias and a disability-related misrepresentation acting on the same underlying variable, producing a compound underestimation that neither bias alone would generate. In each case, the harm occupies a structurally distinct category—one that Crenshaw (1989) identified in legal doctrine over three decades ago—that neither axis of analysis would surface independently.

For researchers, the correction required is not demographic granularity but a fundamental reorientation of study design toward interaction effects. Audit methodology that tests protected attributes sequentially will not detect harms originating in their interaction. Benchmark datasets constructed from majority populations will not yield statistically stable estimates for intersectional subgroups. As Wang et al. (2022) demonstrated empirically, algorithms correcting for single-axis bias frequently fail to improve—and sometimes worsen—outcomes for intersectional subgroups. These are design commitments, not incidental omissions, and correcting them requires treating intersectional subgroup performance as a primary evaluation criterion rather than a secondary sensitivity check.

For practitioners, the implication is that sequential debiasing—correcting for racial bias, then gender bias, in independent interventions applied to the same system—is insufficient by the same logic that makes single-axis auditing insufficient. A system satisfying formal fairness criteria on every individual protected attribute can simultaneously generate compounded harm for every intersectional subgroup, and vendor-supplied single-axis audit results provide no evidence to the contrary. Raghavan et al. (2020) found that none of the eighteen algorithmic hiring vendors they evaluated had documented practices for examining intersectional combinations. That gap is not a technical limitation awaiting a better algorithm; it is an accountability gap that will persist until intersectional evaluation is required rather than voluntary.

For policymakers, the requirement is structural. Non-discrimination frameworks organized around discrete protected classes, and transparency mandates requiring single-attribute disaggregation, reproduce the categorical logic that produces the harms they are meant to address. The parallel between antidiscrimination law's historical failure to recognize intersectional harm and current algorithmic-accountability frameworks' failure to require intersectional evaluation is not incidental—both reflect the assumption that harm concentrated at the intersection of race and gender is reducible to the sum of racial harm and gender harm. That assumption is incorrect as a legal matter, as Crenshaw (1991) established, and this analysis argues it is equally incorrect as a technical matter.

Three questions define the horizon this analysis cannot reach. First, the organizational incentives shaping which intersections receive algorithmic attention—which subgroups enter benchmark datasets, which fairness metrics vendors choose to optimize, which audit findings get published—remain underexamined and require institutional analysis beyond the scope of the cases presented here. Second, the transferability of intersectionality as an analytical framework to other automated-decision domains, including content moderation, credit scoring, and criminal-justice risk assessment, warrants systematic examination rather than assumed generalization from the domains studied here. Third, the accountability mechanisms that would genuinely center intersectionality—rather than appending it as a compliance layer to frameworks designed for single-axis analysis—remain unspecified in both technical and policy literature.

The contribution this analysis makes to the field and study on gender and sexuality is not to add these dimensions to an existing framework. It is to argue that gender identity and sexual orientation are constitutive of how algorithmic systems distribute harm across the full range of intersecting identities—that a cisnormative gender-classification scheme does not affect only trans and nonbinary people, but propagates through every downstream computation in which gender functions as an input, with effects that differ depending on the other social positions a person simultaneously occupies. Buolamwini and Gebru (2018) showed that intersectional error rates in facial classification were not predictable from marginal error rates on race or gender independently. This analysis argues that the same non-additivity characterizes algorithmic harm across hiring, surveillance, and access-gatekeeping contexts—and that the analytical and institutional infrastructure for detecting and remedying it does not yet exist at the scale the problem requires.

REFERENCES

1. Andreassen, R. (2021). Social media surveillance, LGBTQ refugees and asylum: How migration authorities use social media profiles to determine refugees as "genuine" or "fraudulent." First Monday, 26(1). https://doi.org/10.5210/fm.v26i1.10653
2. Alsinai, A., Reygers, M., DiMascolo, L., Kafka, J., Rowhani-Rahbar, A., Adhia, A., Bowen, D., Shanahan, S., Dalve, K., & Ellyson, A. M. (2023). Use of immigration status for coercive control in domestic violence protection orders. Frontiers in Sociology, 8, Article 1146102. https://doi.org/10.3389/fsoc.2023.1146102
3. Baker, R., & Hawn, A. (2022). Algorithmic bias in education. Journal of Learning Analytics, 9(2), 134–151. https://doi.org/10.18608/jla.2022.7847
4. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31
5. Brayne, S. (2020). Predict and surveil: Data, discretion, and the future of policing. Oxford University Press.
6. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Vol. 81, pp. 77–91). Proceedings of Machine Learning Research. https://proceedings.mlr.press/v81/buolamwini18a.html
7. Carpenter, C. S., Eppink, S. T., & Gonzales, G. (2020). Transgender status, gender identity, and socioeconomic outcomes in the United States. ILR Review, 73(3), 573–599. https://doi.org/10.1177/0019793920902776
8. Collins, P. H. (2000). Black feminist thought: Knowledge, consciousness, and the politics of empowerment (2nd ed.). Routledge.

9. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
10. Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. University of Chicago Legal Forum, 1989(1), 139–167. https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8
11. Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review, 43(6), 1241–1299. https://doi.org/10.2307/1229039
12. DeGraffenreid v. General Motors Assembly Division, 413 F. Supp. 142 (E.D. Mo. 1976).
13. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
14. Friedman, C. (2025). Intersecting disability and race implicit attitudes of health care professionals. Disability and Health Journal, 19(1), Article 101924. https://doi.org/10.1016/j.dhjo.2025.101924
15. Gohar, U., & Cheng, L. (2023). A survey on intersectional fairness in machine learning: Notions, mitigation, and challenges. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23) (pp. 6829–6837). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2023/742
16. Greene, J., & Ervin, W. (2024). The cost of crossing gender boundaries: Trans women of color and the racialized workplace gender order. Gender, Work & Organization, 31(6), 2585–2600. https://doi.org/10.1111/gwao.13108
17. Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020). Towards a critical race methodology in algorithmic fairness. In FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 501–512). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372826
18. Idowu, J. A. (2024). Debiasing education algorithms. International Journal of Artificial Intelligence in Education, 34(4), 1510–1540. https://doi.org/10.1007/s40593-023-00389-4
19. Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), Article 88. https://doi.org/10.1145/3274357
20. Lunau, M., & Andreassen, R. (2023). Surveillance practices among migration officers: Online media and LGBTQ+ refugees. International Journal of Cultural Studies, 26(6), 655–671. https://doi.org/10.1177/13678779221140129
21. Ngondwe, P., & Tefera, G. M. (2025). Barriers and facilitators of access to healthcare among immigrants with disabilities: A qualitative meta-synthesis. Healthcare, 13(3), Article 313. https://doi.org/10.3390/healthcare13030313
22. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
23. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
24. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469–481). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372828
25. Riazy, S., Simbeck, K., & Schreck, V. (2020). Fairness in learning analytics: Student at-risk prediction in virtual learning environments. In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) (Vol. 1, pp. 15–25). SCITEPRESS.
26. Sebo, P. (2021). Performance of gender detection tools: A comparative study of name-to-gender inference services. Journal of the Medical Library Association, 109(3), 414–421. https://doi.org/10.5195/jmla.2021.1185
27. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In FAT '19: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598
28. Selod, S. (2018). Forever suspect: Racialized surveillance of Muslim Americans in the war on terror. Rutgers University Press.

29. Sokoloff, N. J., & Dupont, I. (2005). Domestic violence at the intersections of race, class, and gender: Challenges and contributions to understanding violence against marginalized women in diverse communities. Violence Against Women, 11(1), 38–64. https://doi.org/10.1177/1077801204271476
30. Wang, A., Ramaswamy, V. V., & Russakovsky, O. (2022). Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. In FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 336–355). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533101